

Exact Moderate and Large Deviations for Linear Processes

December 6, 2011

Magda Peligrad^a, Hailin Sang^b, Yunda Zhong^c and Wei Biao Wu^c

^a Department of Mathematical Sciences, University of Cincinnati, PO Box 210025, Cincinnati, OH 45221-0025, USA. E-mail address: peligrm@ucmail.uc.edu

^b National Institute of Statistical Sciences, PO Box 14006, Research Triangle Park, NC 27709, USA. E-mail address: sang@niss.org

^c Department of Statistics, University of Chicago, Chicago, IL 60637, USA. E-mail address: ydzhong@galton.uchicago.edu, wbwu@galton.uchicago.edu

Abbreviated Title: Exact Deviations for Linear Processes

Abstract

Large and moderate deviation probabilities play an important role in many applied areas, such as insurance and risk analysis. This paper studies the exact moderate and large deviation asymptotics in non-logarithmic form for linear processes with independent innovations. The linear processes we analyze are general and therefore they include the long memory case. We give an asymptotic representation for probability of the tail of the normalized sums and specify the zones in which it can be approximated either by a standard normal distribution or by the marginal distribution of the innovation process. The results are then applied to regression estimates, moving averages, fractionally integrated processes, linear processes with regularly varying exponents and functions of linear processes. We also consider the computation of value at risk and expected shortfall, fundamental quantities in risk theory and finance.

¹*MSC 2010 subject classification:* 60F10, 62E20

²Key words and phrases: linear process, long memory, moderate deviation, large deviation, zone of normal convergence, non-logarithmic asymptotics.

³This work was supported in part by the Taft Research Center at the University of Cincinnati. In addition, Magda Peligrad was supported by a NSA 2011 grant and a Taft research support grant and Wei Biao Wu by NSF grants DMS-0448704 and DMS-0906073.

1 Introduction and notations

Let $(\xi_i)_{i \in \mathbb{Z}}$ be a sequence of independent and identically distributed centered random variables and c_{ni} a sequence of constants. This paper focuses on the moderate and large deviations in non-logarithmic form for the linear process of the form

$$S_n = \sum_{i=1}^{k_n} c_{ni} \xi_i. \quad (1)$$

This class of linear processes is versatile enough to help analyzing regression estimates, moving averages that include long memory processes, linear processes with regularly varying coefficients, fractionally integrated processes, and functions of linear processes.

Our goal is to find an asymptotic representation for the tail probabilities of the normalized sums defined by (1). Estimations of deviation probabilities occur in a natural way in many applied areas, so for instance, in problems of insurance in the context of large claim insurance.

Specifically, we aim to find a function $N_n(x)$ such that, as $n \rightarrow \infty$,

$$\frac{\mathbb{P}(S_n \geq x\sigma_n)}{N_n(x)} = 1 + o(1), \text{ where } \sigma_n^2 = \|S_n\|_2^2 = \sum_{i=1}^{k_n} c_{ni}^2. \quad (2)$$

If $x \geq 0$ is fixed, then (2) becomes the well-known central limit theorem by letting $N_n(x) = 1 - \Phi(x)$, where $\Phi(x)$ is the standard normal distribution function. In this paper we call $\mathbb{P}(S_n/\sigma_n \geq x)$ the *moderate or large deviation* probabilities depending on the speed of $x \rightarrow \infty$. These tail probabilities of rare events can be very small. Here we call (2) the *exact approximation*, which is more accurate and holds under less restrictive moment conditions than the logarithmic version

$$\frac{\log \mathbb{P}(S_n/\sigma_n \geq x)}{\log N_n(x)} = 1 + o(1). \quad (3)$$

For example, suppose $\mathbb{P}(S_n/\sigma_n \geq x) = 10^{-4}$ and $N_n(x) = 10^{-5}$; then their logarithmic ratio is 0.8, which does not appear to be very different from 1, while the ratio for the exact version (2) is as big as 10. A multiplicative factor of this order can cause substantially different industrial standards in designing projects that can survive natural disasters. The logarithmic version (3) is incapable of effectively characterizing the differences between the tail probabilities.

As early as 1929, Khinchin considered the problem of moderate and large deviation probabilities in non-logarithmic form for independent Bernoulli random variables. The first large deviation probability result appeared in S. Nagaev (1965). A. Nagaev (1969) studied large deviation probabilities of i.i.d. random variables with regularly varying tails. Mikosch and A. Nagaev (1998) applied the large deviation probabilities for heavy-tailed random variables to insurance mathematics. The review work on this topic can be found in S. Nagaev (1979)

and Rozovski (1993). Rubin and Sethuraman (1965), Slastnikov (1978) and Frolov (2005) considered the moderate or large deviations for arrays of independent random variables. S. Nagaev (1979) presented the following very useful result: in (1) assume $k_n = n$, $c_{ni} \equiv 1$, and that ξ_i has a regularly varying right tail. i.e.

$$\mathbb{P}(\xi_0 \geq x) = \frac{h(x)}{x^t}(1 + o(1)) \text{ as } x \rightarrow \infty \text{ for some } t > 2, \quad (4)$$

where $h(x)$ is a slowly varying function (Bingham, Goldie and Teugels, 1987). Namely, $\lim_{x \rightarrow \infty} h(\lambda x)/h(x) = 1$ for all $\lambda > 0$. Note that $\sigma_n = \sqrt{n}$. If in addition, for some $p > 2$, ξ_0 has absolute moment of order p , then

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \geq x\sigma_n\right) = (1 - \Phi(x))(1 + o(1)) + n\mathbb{P}(\xi_0 \geq x\sigma_n)(1 + o(1)) \quad (5)$$

for $n \rightarrow \infty$ and $x \geq 1$. Note that (5) implies (2) with

$$N_n(x) = (1 - \Phi(x)) + n\mathbb{P}(\xi_0 \geq x\sigma_n). \quad (6)$$

Hence if $1 - \Phi(x) = o[n\mathbb{P}(\xi_0 \geq x\sigma_n)]$ (resp. $n\mathbb{P}(\xi_0 \geq x\sigma_n) = o(1 - \Phi(x))$), then in (2) we can also choose $N_n(x) = 1 - \Phi(x)$ (resp. $N_n(x) = n\mathbb{P}(\xi_0 \geq x\sigma_n)$).

The study of moderate and large deviation probabilities in non-logarithmic form for dependent random variables is still in its initial stage. Ghosh (1974) considered moderate deviations for m -dependent random variables. Chen (2001) obtained moderate deviation result for Markov processes. Grama (1997) and Grama and Haeusler (2006) investigated the martingale case. Wu and Zhao (2008) studied moderate deviations for stationary processes which applies to many time series models. However the result in the latter paper can only be applied to linear processes with short memory and their transformations.

For analyzing linear processes with long memory and for obtaining other interesting applications, we study processes of type (1). Under mild conditions on the coefficients, we shall point out the zones in which the deviation probabilities can be approximated either by a standard normal distribution or by using the distribution of ξ_0 . Our main result is that (5) holds in our case with

$$N_n(x) = (1 - \Phi(x)) + \sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \geq x\sigma_n).$$

The paper has the following structure. Section 2 presents a general moderate and large deviation result and various applications. Section 3 illustrates the results of a numeric study. In Section 4 we prove the results. In the Appendix we give some auxiliary results and we also mention some known facts needed for the proofs.

Before stating our results we introduce the notations used throughout this paper: $a_n \sim b_n$ means that $\lim_{n \rightarrow \infty} a_n/b_n = 1$, $a_n = O(b_n)$ and also $a_n \ll b_n$ means $\limsup_{n \rightarrow \infty} a_n/b_n < \infty$; $a_n = o(b_n)$ if $\lim_{n \rightarrow \infty} a_n/b_n = 0$. By $\|X\|_p$ we denote $(\mathbb{E}|X|^p)^{1/p}$. The notation $l(\cdot)$, $h(\cdot)$ and $\ell(\cdot)$ denote slowly varying functions.

2 Main Results

It is convenient to normalize by the variance and throughout the paper, we assume that:

Condition A. $(\xi_i)_{i \in \mathbb{Z}}$, are i.i.d. centered random variables with finite second moment, $\mathbb{E}(\xi_0^2) = 1$.

2.1 General linear processes

Our first results apply to general linear processes of type (1) with i.i.d. innovations. For $c_{ni} > 0$ and $t > 0$ we define

$$B_{nt} = \sum_{i=1}^{k_n} c_{ni}^t, \quad (7)$$

$$\sigma_n^2 = \text{var}(S_n) = B_{n2}, \quad (8)$$

and

$$D_{nt} = B_{n2}^{-t/2} B_{nt}. \quad (9)$$

The basic assumption in all our results is the uniformly asymptotically negligibility of the variance of individual summands, namely

$$\max_{1 \leq i \leq k_n} c_{ni}^2 / \sigma_n^2 \rightarrow 0. \quad (10)$$

Our first theorem extends Nagaev's result in (5) to general linear processes.

Theorem 2.1 *Assume that $(\xi_i)_{i \in \mathbb{Z}}$ satisfies Condition A, and for a certain $t > 2$ it satisfies the right tail condition (4). Moreover, for a certain $p > 2$, $\|\xi_0\|_p < \infty$. Assume also that $c_{ni} > 0$ and (10) is satisfied. Then, as $n \rightarrow \infty$,*

$$\mathbb{P}(S_n \geq x\sigma_n) = (1 + o(1)) \sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \geq x\sigma_n) + (1 - \Phi(x))(1 + o(1)) \quad (11)$$

holds for all $x > 0$.

Corollary 2.1 *Under the conditions of Theorem 2.1 for $x \geq a(\ln D_{nt}^{-1})^{1/2}$ with $a > 2^{1/2}$ we have*

$$\mathbb{P}(S_n \geq x\sigma_n) = (1 + o(1)) \sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \geq x\sigma_n) \text{ as } n \rightarrow \infty. \quad (12)$$

On the other hand, if $0 < x \leq b(\ln D_{nt}^{-1})^{1/2}$ with $b < 2^{1/2}$, we have

$$\mathbb{P}(S_n \geq x\sigma_n) = (1 - \Phi(x))(1 + o(1)) \text{ as } n \rightarrow \infty. \quad (13)$$

Remark 2.1 Notice that (12) suggests that if $x \geq a(\ln D_{nt}^{-1})^{1/2}$ where $a > 2^{1/2}$, then (2) holds with

$$N_n(x) = \sum_{i=1}^{k_n} \frac{h(x\sigma_n/c_{ni})}{(x\sigma_n/c_{ni})^t}.$$

For the special case in which $\lim_{x \rightarrow \infty} h(x) \rightarrow h_0 > 0$, we can also choose

$$N_n(x) = \frac{h_0}{x^t} \sum_{i=1}^{k_n} \left(\frac{c_{ni}}{\sigma_n}\right)^t = \frac{h_0}{x^t} D_{nt}. \quad (14)$$

As a consequence to Theorem 2.1 we have

Corollary 2.2 Assume that $(\xi_i)_{i \in \mathbb{Z}}$ satisfies Condition A, and for a certain $t > 2$ it satisfies

$$\mathbb{P}(|\xi_0| \geq x) = \frac{h(x)}{x^t} (1 + o(1)) \text{ as } x \rightarrow \infty. \quad (15)$$

Assume also that $c_{ni} > 0$ and (10) is satisfied. Then the conclusions of Theorem 2.1 and Corollary 2.1 are valid.

Notice that (12) and (13) assert different approximations for the tail probability $\mathbb{P}(S_n \geq x\sigma_n)$: moderate behavior for x smaller than a threshold, when we can approximate this probability by using a normal distribution. On the other hand we have a large deviation type of behavior for x larger than another threshold, where S_n exceeds a level because essentially one of the summands is large.

The proofs of these results are based on a separate study of the behaviors of type (12) or (13), which is of independent interest. As a matter of fact, we shall see in the next two theorems that a result similar to (12) holds without the assumption of the finite moment of order p while the moderate deviation (13) does not require a regularly varying right tails.

Theorem 2.2 Assume that $(\xi_i)_{i \in \mathbb{Z}}$ satisfies Condition A, and for a certain $t > 2$ it satisfies (4). Let $c_{ni} > 0$ be a sequence of constants satisfying (10). Then, for $x \geq C_t(\ln D_{nt}^{-1})^{1/2}$ with $C_t > e^{t/2}(t+2)/\sqrt{2}$ the large deviation result (12) holds.

As a counterpart to this result we shall formulate now the moderate deviation bound. Recall that $\Phi(x)$ is the standard normal distribution function. Notice that D_{nt} is, up to a factor involving the moments of the innovations, the Lyapunov's proportion.

Theorem 2.3 Assume that $(\xi_i)_{i \in \mathbb{Z}}$ satisfies Condition A and for a certain $p > 2$, $\|\xi_0\|_p < \infty$. Assume that (10) is satisfied. If $x^2 \leq 2\ln(D_{np}^{-1})$ then the moderate deviation result (13) holds.

2.2 Applications to linear regression estimates

Many random evolutions and also statistical procedures, such as estimation of regression coefficients, produce linear statistics of type (1). See for instance Chapter 9 in Beran (1994), for the case of parametric regression, or the paper by Robinson (1997), where kernel estimators are used for nonparametric regression. Here we consider the simple parametric regression model $Y_i = \beta\alpha_i + \xi_i$, where (ξ_i) is an i.i.d. sequence of errors, (α_i) is a sequence of positive real numbers and β is the parameter of interest. The least squares estimator $\hat{\beta}_n$ of β , based on a sample of size n , satisfies

$$S_n := \hat{\beta}_n - \beta = \frac{1}{\sum_{i=1}^n \alpha_i^2} \sum_{i=1}^n \alpha_i \xi_i, \quad (16)$$

so, the representation of type (1) holds with $c_{ni} = \alpha_i / (\sum_{i=1}^n \alpha_i^2)$. Denote $A_{nt} = \sum_{i=1}^n \alpha_i^t$. Notice that $\text{var}(S_n) = 1/A_{n2}$.

Assume

$$\lim_{n \rightarrow \infty} A_{n2}^{-1} \max_{1 \leq i \leq n} \alpha_i^2 = 0. \quad (17)$$

As an immediate consequence of Theorem 2.1, we obtain:

Corollary 2.3 (i) Assume that $(\xi_i)_{i \in \mathbb{Z}}$ satisfies the conditions in Theorem 2.1. Under assumption (17), for $x > 0$, we have

$$\begin{aligned} \mathbb{P}(\hat{\beta}_n - \beta \geq x/A_{n2}^{1/2}) = \\ (1 + o(1)) \sum_{i=1}^n \mathbb{P}(\xi_i \geq x A_{n2}^{1/2} / \alpha_i) + (1 + o(1))(1 - \Phi(x)). \end{aligned}$$

(ii) If $x > 0$ and $x^2 \leq 2 \ln(A_{n2}^{t/2}/A_{nt})$, we have

$$\mathbb{P}(\hat{\beta}_n - \beta \geq x/A_{n2}^{1/2}) = (1 + o(1))(1 - \Phi(x)).$$

(iii) If $x > 0$ and $x^2 \geq C_t^2 \ln(A_{n2}^{t/2}/A_{nt})$ with $C_t^2 > 2$ then

$$\mathbb{P}(\hat{\beta}_n - \beta \geq x/A_{n2}^{1/2}) = (1 + o(1)) \sum_{i=1}^n \mathbb{P}(\xi_i \geq x A_{n2}^{1/2} / \alpha_i).$$

Similar results as in Theorem 2.2 and Theorem 2.3 can also be easily formulated.

Theorems 2.1, 2.2 and 2.3 are also applicable to the nonlinear regression model $y_i = g(x_i) + \xi_i$, $1 \leq i \leq n$, where $g(x)$ is an unknown function and ξ_i is the noise. Let x_i be the deterministic design points. Then the Nadaraya-Watson estimate \hat{g}_n satisfies

$$\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x) = \sum_{i=1}^n c_{ni}(x) \xi_i$$

with

$$c_{ni}(x) = K\left(\frac{x_i - x}{h_n}\right) / \sum_{i=1}^n K\left(\frac{x_i - x}{h_n}\right),$$

where K is a kernel function and $h_n > 0$ is a sequence of bandwidths converging to 0, and therefore is of the type (1).

2.3 Application to moving averages

We now consider the sum $S_n = \sum_{k=1}^n X_k$, where

$$X_k = \sum_{j=-\infty}^{\infty} a_{k-j} \xi_j. \quad (18)$$

We assume that $\sum_{i \in \mathbb{Z}} a_i^2 < \infty$, which is the necessary and sufficient condition for the existence of X_1 . Observe that $S_n = \sum_{i=-\infty}^{\infty} b_{ni} \xi_i$ is of form (1) with

$$b_{nj} = a_{1-j} + \cdots + a_{n-j}. \quad (19)$$

Define D_{nt} by (9) with $c_{ni} = b_{ni}$. We know from Peligrad and Utev (1997) that under the assumption $\sigma_n^2 \rightarrow \infty$ we have

$$\sigma_n^{-2} \sup_i b_{ni}^2 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Therefore condition (10) is automatically satisfied and as a corollary of Theorems 2.1, 2.2, and 2.3 we obtain:

Corollary 2.4 *Assume that $(X_n)_{n \geq 1}$ is defined by (18) and $\sigma_n^2 \rightarrow \infty$.*

- (i) Assume that $(\xi_i)_{i \in \mathbb{Z}}$ satisfies the conditions of Theorem 2.1 and $b_{nk} \geq 0$; then (11) holds.*
- (ii) Let $(\xi_i)_{i \in \mathbb{Z}}$ be as in Theorem 2.2. Assume $b_{nk} \geq 0$; then the large deviation result (12) holds.*
- (iii) Assume $(\xi_i)_{i \in \mathbb{Z}}$ is as in Theorem 2.3; then the moderate deviation result (13) is valid.*

Notice that this corollary applies to general linear processes including the long memory processes with $\sum_i |a_i| = \infty$. Asymptotic properties for long memory processes can be quite different from those of processes with short memory, partially because the variance of the partial sum goes to infinity at an order different than n ; see for example, Ho and Hsing (1997), Robinson (2003), Doukhan, Oppenheim and Taqqu (2003) among others. Hall (1992) gave a Berry-Esseen bound for the convergence rate in the central limit theorem.

We shall apply now this corollary to the important particular case of causal long-memory processes with

$$a_i = l(i+1)(1+i)^{-r}, \quad i \geq 0, \text{ with } 1/2 < r < 1, \text{ and } a_i = 0 \text{ in rest.} \quad (20)$$

Here $l(\cdot)$ is a slowly varying function where the results can be given in a more precise form. Notice that in this particular case

$$X_k = \sum_{j=-\infty}^k a_{k-j} \xi_j.$$

Let $a_0 = 1$. This case of long memory linear processes covers the well-known fractional ARIMA processes (cf. Granger and Joyeux, 1980, Hosking, 1981), which plays an important role in financial time series modeling and application. As a special case, let $0 < d < 1/2$ and B be the backward shift operator with $B\varepsilon_k = \varepsilon_{k-1}$ and consider

$$X_k = (1 - B)^{-d} \xi_k = \sum_{i \geq 0} a_i \xi_{k-i}, \text{ where } a_i = \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}.$$

For this example we have $\lim_{n \rightarrow \infty} a_n/n^{d-1} = 1/\Gamma(d)$. Notice that these processes have long memory because $\sum_{j \geq 0} |a_j| = \infty$.

Corollary 2.5 *Assume (20). If $(\xi_i)_{i \in \mathbb{Z}}$ satisfies the conditions of Theorem 2.1 then (11) holds. In particular (12) holds for $x \geq c_1(\ln n)^{1/2}$ with $c_1 > (t-2)^{1/2}$ while (13) holds, provided $0 < x \leq c_2(\ln n)^{1/2}$ with $c_2 < (t-2)^{1/2}$.*

For this case Theorems 2.2 and 2.3 give:

Corollary 2.6 (i) *Let $(\xi_i)_{i \in \mathbb{Z}}$ be as in Theorem 2.2. Then (12) holds for $x > c_1(\ln n)^{1/2}$ with $c_1 > (t-2)^{1/2}e^{t/2}(t+2)/2$.*
(ii) *Assume $(\xi_i)_{i \in \mathbb{Z}}$ is as in Theorem 2.3. Then (13) holds, provided $x^2 \leq (p-2)(\ln n)$.*

2.4 Application to risk measures

In risk theory and finance, value at risk (VaR) and expected shortfall (ES) play a fundamental role; see Jorion (2006), Holton (2003), McNeil et al (2005), Acerbi and Tasche (2002) among others. Mathematically, they are equivalent to quantiles and tail conditional expectations. In practice one is most interested in their extremal behavior which corresponds to tail quantiles. Despite their importance, however, their computation can be quite difficult and the related asymptotic justification is far from being trivial.

Here we shall apply Theorem 2.1 and provide approximate formulae for extremal quantiles and tail conditional expectations for S_n . Under the assumption $\lim_{x \rightarrow \infty} h(x) \rightarrow h_0 > 0$, by (14) and Theorem 2.1,

$$\mathbb{P}(S_n \geq x\sigma_n) = (1 + o(1)) \frac{h_0}{x^t} D_{nt} + (1 - \Phi(x))(1 + o(1)).$$

Given the tail probability $\alpha \in (0, 1)$, let $q_{\alpha, n}$ be the upper α -th quantile of S_n . Namely $\mathbb{P}(S_n \geq q_{\alpha, n}) = \alpha$. Elementary calculations show that $q_{\alpha, n}$ can be

approximated by $x_\alpha \sigma_n$ in the sense that $\lim_{n \rightarrow \infty} x_\alpha \sigma_n / q_{\alpha,n} = 1$, where $x = x_\alpha$ is the solution to the equation

$$\frac{h_0}{x^t} D_{nt} + (1 - \Phi(x)) = \alpha.$$

In particular, if $\alpha \leq h_0 D_{nt} (a^2 \ln D_{nt}^{-1})^{-t/2}$ with $a > 2^{1/2}$, then, by Corollary 2.1, we can approximate $q_{\alpha,n}$ by $(h_0 D_{nt} / \alpha)^{1/t} \sigma_n = (B_{nt} h_0 / \alpha)^{1/t}$. The approximation is understood in the sense that $(B_{nt} h_0 / \alpha)^{1/t} / q_{\alpha,n} \rightarrow 1$ as $n \rightarrow \infty$, and the tail conditional expectation or expected shortfall is computed as

$$\begin{aligned} \mathbb{E}(S_n | S_n \geq q_{\alpha,n}) &= \frac{q_{\alpha,n} \mathbb{P}(S_n \geq q_{\alpha,n}) + \int_{q_{\alpha,n}}^{\infty} \mathbb{P}(S_n \geq w) dw}{\mathbb{P}(S_n \geq q_{\alpha,n})} \\ &\sim q_{\alpha,n} + \frac{q_{\alpha,n}}{t-1} = \frac{t q_{\alpha,n}}{t-1} \sim B_{nt}^{1/t} \frac{t (h_0 / \alpha)^{1/t}}{t-1}. \end{aligned}$$

We emphasize that, without the exact moderate deviation principle of Theorem 2.1, the validity of the above equivalence cannot be guaranteed. To the best of our knowledge, our example is one of the very few cases that one can obtain an explicit asymptotic expression for VaR and ES for sums of dependent random variables.

2.5 Functionals of linear processes

In this subsection we shall use the result from the point (ii) of Corollary 2.6 to study the moderate deviation for nonlinear transformations of linear processes. Let K be a transformation which is measurable and $\mathbb{E}K(X_0) = 0$. Let

$$H_n = \sum_{i=1}^n K(X_i).$$

For example, if $K(X_0) = I(X_0 \leq \tau) - \mathbb{P}(X_0 \leq \tau)$, then H_n/n becomes the empirical process. If X_i is a short memory linear process, namely a_i are absolutely summable and their sum is different of 0, then we can apply the moderate deviation principle in Wu and Zhao (2008). However, the result in the latter paper is not applicable for long-range dependent processes. Despite its importance in risk analysis, the problem of moderate deviation under strong dependence has been rarely studied in the literature.

Here we shall establish such a principle in the context of nonlinear transforms of linear processes. First, we introduce some necessary notation for this section. Let $\mathcal{F}_n = (\dots, \xi_{n-1}, \xi_n)$ be the shift process and define the projection operator $\mathcal{P}_i \cdot = \mathbb{E}(\cdot | \mathcal{F}_i) - \mathbb{E}(\cdot | \mathcal{F}_{i-1})$. Denote the truncated processes $X_{n,k} = \mathbb{E}(X_n | \mathcal{F}_k)$. Now define the functions $K_n(w) = \mathbb{E}[K(w + X_n - X_{n,0})]$ and $K_\infty(w) = \mathbb{E}[K(w + X_n)]$. We consider transformations K with $\kappa := K'_\infty(0) \neq 0$. Define

$$S_{n,1} = \sum_{i=1}^n [K(X_i) - \kappa X_i] = H_n - \kappa S_n, \text{ where } S_n = \sum_{i=1}^n X_i.$$

Then $H_n = \kappa S_n + S_{n,1}$. For a function g , let $g(w; \lambda) = \sup_{|y| \leq \lambda} |g(w+y)|$ be the local maximal function. Denote the collection of functions with second order partial derivatives by $\mathbb{C}^2(\mathbb{R})$. We need the following regularity condition.

Condition B. Let $2 \leq q < p \leq 2q$ and assume $\|\xi_0\|_p < \infty$. Assume $K_n \in C^2(\mathbb{R})$ for all large n and that for some $\lambda > 0$,

$$\sum_{i=0}^2 \|K_{n-1}^{(i)}(X_{n,0}; \lambda)\|_q + \|\xi_1\|^{p/q} K_{n-1}(X_{n,1})\|_q + \|\xi_1 K'_{n-1}(X_{n,1})\|_q = O(1).$$

A version of Condition B with $q = 2$ is used in Wu (2006). We shall establish the following moderate deviation result. For $1/2 < r < 1$ and $1/2 \leq v < 1$ define

$$\begin{aligned} \chi(v, r) &= v \max(r - r/v, 1/2 - r, r - 1), \\ \omega(r) &= \arg\min_{1/2 \leq v < 1} \chi(v, r) \text{ and } \rho(r) = -\chi(\omega(r), r). \end{aligned}$$

Theorem 2.4 *Assume that Condition B holds with $q = p\omega(r)$ and the conditions of Corollary 2.5 (ii) are satisfied. Let c be such that $0 < c \leq p - 2$ and $c < 2p\rho(r)$. Then if $x \leq c \ln n$, we have*

$$\mathbb{P}(H_n \geq |\kappa| \sigma_n x) = (1 - \Phi(x))(1 + o(1)) \text{ as } n \rightarrow \infty. \quad (21)$$

Remark 2.2 *As mentioned in the proof of Theorem 2.4 in Section 4.7, (21) is still valid if the normalizing constant $|\kappa| \sigma_n$ therein is replaced by $\sqrt{\text{var}(H_n)}$.*

Remark 2.3 *Theorem 2.4 only asserts a moderate deviation with the Gaussian range. It is unclear whether the approximation of type (12) holds. We pose it as an open problem.*

Remark 2.4 *An explicit form for $\omega(r)$ can be obtained. If $r \geq 3/4$, then $\omega(r) = r$. If $r < 3/4$, then $\omega(r) = r/(2r - 1/2)$. If $2p\rho(r) \geq p - 2$, then the moderate deviation in (21) has the same range as for S_n . The latter happens, for example, if $r = 3/4$ and $2 < p < 16/5$, since in this case $2p\rho(3/4) \geq p - 2$.*

Example 2.1 *As an application to empirical processes, let $K(X) = I(X \leq \tau) - \mathbb{P}(X \leq \tau)$, where $\tau \in \mathbb{R}$ is fixed. Let $X_n = \xi_n + \sum_{i=1}^{\infty} a_i \xi_{n-i} =: \xi_n + Y_{n-1}$, where $\|\xi_0\|_p < \infty$, $p > 2$, and its density function f_ξ satisfies*

$$\sup_u [f_\xi(u) + |f'_\xi(u)|] < \infty. \quad (22)$$

Then $K_1(w) = F_\xi(\tau - w) - F_X(\tau)$, where F_ξ is the distribution function of ξ_i . Under (22), we clearly have $\sup_w [|K'_1(w)| + |K''_1(w)|] < \infty$. Observe that we have the identity: for $n \geq 1$,

$$K_n(w) = \mathbb{E} K_1(w + a_1 \xi_{n-1} + a_2 \xi_{n-2} + \dots + a_{n-1} \xi_1).$$

Hence $\sup_n \sup_w [|K'_n(w)| + |K''_n(w)|] < \infty$. So Condition B holds for any λ since $\xi_n \in L^p$, $p > 2$.

3 A Numerical Study

In this section we shall design a numeric study of the accuracy of the large deviation (12), normal approximation (13) and also the estimate (11) on a finite simple. In particular, we shall study the accuracy of the approximations in Corollary 2.5. In general it is very difficult to calculate tail probabilities by simulation, especially if they are small. One may need to carry out astronomically large amount of computations to obtain reasonably well approximations.

Here we shall approach the problem from a different angle. We let $X_j = \sum_{i=1}^{\infty} a_i \xi_{j-i}$, where $\xi_i, i \in \mathbb{Z}$, have Student's t-distribution with degree of freedom $\nu = 3$, and $a_i = i^{-0.9}$. Let $S_n = \sum_{i=1}^n X_i$ with $n = 300$. Note that the characteristic function of ξ_i is

$$\varphi(t) = \frac{(\sqrt{\nu}|t|)^{\nu/2} K_{\nu/2}(\sqrt{\nu}|t|)}{\Gamma(\nu/2) 2^{\nu/2-1}}, \quad (23)$$

where $K_{\nu/2}$ is the Bessel function (see Hurst (1995)). Then the characteristic function of S_n is

$$\varphi_{S_n}(t) = \prod_{j \in \mathbb{Z}} \varphi(b_{nj}t)$$

and by the inversion formula,

$$\mathbb{P}(S_n \leq x) - \mathbb{P}(S_n \leq x') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\sqrt{-1}yx} - e^{\sqrt{-1}yx'}}{\sqrt{-1}y} \varphi_{S_n}(y) dy.$$

In the above equation let $x' = 0$. Since ξ_j is symmetric, $\mathbb{P}(S_n \leq 0) = 1/2$. In our numeric study we shall use (23) to compute the probability $\mathbb{P}(S_n > x)$.

In Figure 3 we report the ratios $R(x) := \sum_i \mathbb{P}(b_{ni}\xi_0 \geq x)/\mathbb{P}(S_n > x)$ and $g(x) := (1 - \Phi(x/\sigma_n))/\mathbb{P}(S_n > x)$; see (12) with $c_{ni} = b_{ni}$. We can interpret $R(x)$ (resp. $g(x)$) as tail (resp. Gaussian) approximation. As expected from Corollary 2.5, the Gaussian approximation is better if x is small, while the tail probability $R(x)$ approximation is better when x is big. In the intermediate region we approximate by their sum.

4 Proofs

4.1 Preliminary approximations

Let $(X_{ni})_{1 \leq i \leq k_n}$ be a triangular array of independent random variables. We shall approximate here the tail distribution of partial sums by the tail of the sums of truncated random variables and a term involving the tail probabilities of individual summands. We implement the following notations:

$$S_n = \sum_{i=1}^{k_n} X_{ni}, \quad S_n(j) = \sum_{i \neq j}^{k_n} X_{ni}$$

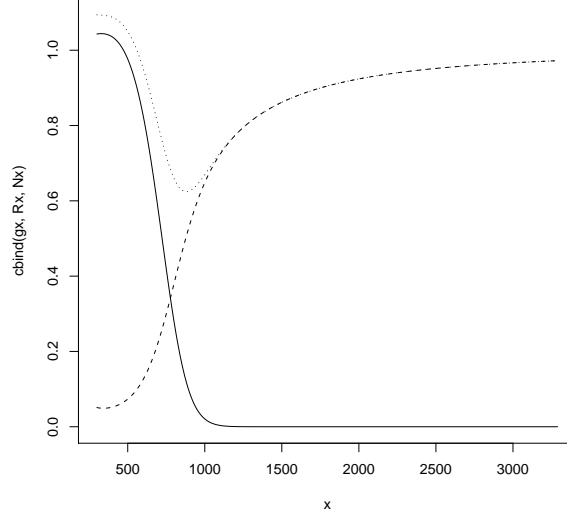


Fig. 1. Tail approximation $R(x)$ (dashed curve), Gaussian approximation $g(x)$ (solid curve) and their sum (dotted curve) for long-memory processes with Student $t(3)$ innovations.

and for $x > 0$ and $\varepsilon > 0$ we set

$$X_{ni}^{(\varepsilon x)} = X_{ni}I(X_{ni} < \varepsilon x), S_n^{(\varepsilon x)} = \sum_{i=1}^{k_n} X_{ni}^{(\varepsilon x)} \text{ and } S_n^{(\varepsilon x)}(j) = \sum_{i \neq j}^{k_n} X_{ni}^{(\varepsilon x)}. \quad (24)$$

We shall prove the following key lemma that will be further exploited to approximate the tail distribution of $\mathbb{P}(S_n \geq x)$ in terms of the sum of the truncated random variables and the tail distributions of the individual summands.

Lemma 4.1 *For any $0 < \eta < 1$, and $\varepsilon > 0$ such that $1 - \eta > \varepsilon$ we have*

$$\begin{aligned} & |\mathbb{P}(S_n \geq x) - \mathbb{P}(S_n^{(\varepsilon x)} \geq x) - \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq (1 - \eta)x)| \leq \\ & 4\left(\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x)\right)^2 + 3\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x)(\mathbb{P}(|S_n(j)| > \eta x) \\ & + \sum_{j=1}^{k_n} \mathbb{P}((1 - \eta)x \leq X_{nj} < (1 + \eta)x). \end{aligned}$$

Proof. We start to estimate $S_n \geq x$ by using the decomposition according to $\max_{i \neq j} X_{ni} < \varepsilon x$ or $\max_{i \neq j} X_{ni} \geq \varepsilon x$, and the last one can happen if exactly one of the variables is larger than εx or at least two variables exceed εx . Formally,

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &= \sum_{j=1}^{k_n} \mathbb{P}(S_n \geq x, X_{nj} \geq \varepsilon x, \max_{i \neq j} X_{ni} < \varepsilon x) \\
&+ \sum_{i=1}^{k_n-1} \sum_{j=i+1}^{k_n} \mathbb{P}(S_n \geq x, X_{nj} \geq \varepsilon x, X_{ni} \geq \varepsilon x) + \mathbb{P}(S_n \geq x, \max_{1 \leq i \leq k_n} X_{ni} < \varepsilon x) \\
&= A + B + C = \sum_{j=1}^{k_n} A_j + B + C.
\end{aligned}$$

The term B can be easily majorated by

$$B \leq \sum_{i=1}^{k_n-1} \sum_{j=i+1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) \mathbb{P}(X_{ni} \geq \varepsilon x) \leq \left(\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) \right)^2.$$

We analyze now the first term. We introduce a new parameter $\eta > 0$. Since for any two events A and B we have $|P(A) - P(B)| \leq P(AB') + P(A'B)$, (here the prime stays for the complement), for each j we have

$$\begin{aligned}
|A_j - \mathbb{P}(X_{nj} \geq (1 - \eta)x)| &\leq \mathbb{P}(S_n \geq x, X_{nj} \geq \varepsilon x, X_{nj} < (1 - \eta)x) \\
&+ \mathbb{P}(X_{nj} \geq (1 - \eta)x, S_n < x) + \mathbb{P}(X_{nj} \geq (1 - \eta)x, X_{nj} < \varepsilon x) \\
&+ \mathbb{P}(X_{nj} \geq (1 - \eta)x, \max_{i \neq j} X_{ni} \geq \varepsilon x) = I + II + III + IV.
\end{aligned}$$

We treat each term separately. By independence and since $S_n \geq x$ and $X_{nj} < (1 - \eta)x$ imply $S_n(j) \geq \eta x$, we derive

$$I \leq \mathbb{P}(X_{nj} \geq \varepsilon x) \mathbb{P}(S_n(j) \geq \eta x).$$

The second term is treated in the following way:

$$\begin{aligned}
II &\leq \mathbb{P}((1 - \eta)x \leq X_{nj} < (1 + \eta)x) + \mathbb{P}(X_{nj} \geq (1 + \eta)x, S_n < x) \\
&\leq \mathbb{P}((1 - \eta)x \leq X_{nj} < (1 + \eta)x) + \mathbb{P}(X_{nj} \geq (1 + \eta)x) \mathbb{P}(-S_n(j) \geq x).
\end{aligned}$$

Since $1 - \eta > \varepsilon$ the third term is: $III = 0$. By independence, the forth term is

$$IV = \mathbb{P}(X_{nj} \geq (1 - \eta)x) \mathbb{P}(\max_{i \neq j} X_{ni} \geq \varepsilon x).$$

Overall, by the previous estimates and because $1 - \eta > \varepsilon$, we obtain

$$\begin{aligned}
|A - \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq (1 - \eta)x)| &\leq 2 \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) (\mathbb{P}(|S_n(j)| > \eta x) \\
&+ \left(\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) \right)^2 + \sum_{j=1}^{k_n} \mathbb{P}((1 - \eta)x \leq X_{nj} < (1 + \eta)x).
\end{aligned}$$

It remains to analyze the last term, C . Notice that

$$\begin{aligned} |C - \mathbb{P}(S_n^{(\varepsilon x)} \geq x)| &= \mathbb{P}(S_n^{(\varepsilon x)} \geq x) - \mathbb{P}(S_n^{(\varepsilon x)} \geq x, \max_{1 \leq i \leq k_n} X_{ni} < \varepsilon x) \\ &= \mathbb{P}(S_n^{(\varepsilon x)} \geq x, \max_{1 \leq i \leq k_n} X_{ni} \geq \varepsilon x). \end{aligned}$$

Now we treat this term by the same arguments we have already used, by dividing the maximum in two parts:

$$\begin{aligned} \mathbb{P}(S_n^{(\varepsilon x)} \geq x, \max_{1 \leq i \leq k_n} X_{ni} \geq \varepsilon x) &= \sum_{j=1}^{k_n} \mathbb{P}(S_n^{(\varepsilon x)} \geq x, X_{nj} \geq \varepsilon x, \max_{i \neq j} X_{ni} < \varepsilon x) \\ &+ \sum_{i=1}^{k_n-1} \sum_{j=i+1}^{k_n} \mathbb{P}(S_n^{(\varepsilon x)} \geq x, X_{nj} \geq \varepsilon x, X_{ni} \geq \varepsilon x) = \sum_{j=1}^{k_n} F_j + G. \end{aligned}$$

The last term, G is majorated exactly as B . As for the first term, we notice that because $X_{nj} \geq \varepsilon x$ the term $X_{nj}^{(\varepsilon x)}$ does not appear in the sum, and by independence we obtain

$$\begin{aligned} F_j &= \mathbb{P}(S_n^{(\varepsilon x)}(j) \geq x, X_{nj} \geq \varepsilon x, \max_{i \neq j} X_{ni} < \varepsilon x) \\ &\leq \mathbb{P}(S_n^{(\varepsilon x)}(j) \geq x) \mathbb{P}(X_{nj} \geq \varepsilon x). \end{aligned}$$

Now, clearly we have

$$\begin{aligned} \mathbb{P}(S_n^{(\varepsilon x)}(j) \geq x) &\leq \mathbb{P}(\max_i X_{ni} \geq \varepsilon x) + \mathbb{P}(S_n^{(\varepsilon x)}(j) \geq x, \max_i X_{ni} < \varepsilon x) \\ &= \mathbb{P}(\max_i X_{ni} \geq \varepsilon x) + \mathbb{P}(S_n(j) \geq x, \max_i X_{ni} < \varepsilon x), \end{aligned}$$

implying that

$$\sum_{j=1}^{k_n} F_j \leq \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) (\mathbb{P}(\max_i X_{ni} \geq \varepsilon x) + \mathbb{P}(S_n(j) \geq x)).$$

Overall,

$$|C - \mathbb{P}(S_n^{(\varepsilon x)} \geq x)| \leq 2 \left(\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) \right)^2 + \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x) \mathbb{P}(S_n(j) \geq x).$$

By gathering all the information above and taking into account that

$$\begin{aligned} |\mathbb{P}(S_n \geq x) - \mathbb{P}(S_n^{(\varepsilon x)} \geq x) - \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq (1-\eta)x)| &\leq \\ |A - \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq (1-\eta)x)| &+ |C - \mathbb{P}(S_n^{(\varepsilon x)} \geq x)| + |B|, \end{aligned}$$

the lemma is established. \diamond

If S_n is stochastically bounded, i.e. $\lim_{K \rightarrow \infty} \sup_n \mathbb{P}(|S_n| > K) = 0$, the approximation in Lemma 4.1 have a simple asymptotic form.

Proposition 4.1 *Assume S_n is stochastically bounded, the variables are centered, and $x_n \rightarrow \infty$. Then for any $0 < \eta < 1$, and $\varepsilon > 0$ such that $1 - \eta > \varepsilon$,*

$$\begin{aligned} \mathbb{P}(S_n \geq x_n) &= \mathbb{P}(S_n^{(\varepsilon x_n)} \geq x_n) + \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq (1 - \eta)x_n) \\ &+ o\left(\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x_n)\right) + \sum_{j=1}^{k_n} \mathbb{P}((1 - \eta)x_n \leq X_{nj} < (1 + \eta)x_n). \end{aligned}$$

Proof. We just notice that for independent centered random variables, if S_n is stochastically bounded, by Lévy inequality (Inequality 1.1.3 in de la Peña and Giné 1999), we have $\max_{1 \leq i \leq k_n} |X_{ni}|$ is stochastically bounded too. By taking into account that $|S_n(j)| \leq |S_n| + \max_{1 \leq i \leq k_n} |X_{ni}|$, we obtain

$$\begin{aligned} \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x_n) \mathbb{P}(|S_n(j)| \geq \eta x_n) &\leq \max_{1 \leq j \leq k_n} \mathbb{P}(|S_n(j)| \geq \eta x_n) \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x_n) \leq \\ &\left(\mathbb{P}(|S_n| \geq \eta x_n/2) + \mathbb{P}\left(\max_{1 \leq i \leq k_n} |X_{ni}| \geq \eta x_n/2\right) \right) \sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x_n) \\ &= o\left(\sum_{j=1}^{k_n} \mathbb{P}(X_{nj} \geq \varepsilon x_n)\right) \text{ as } n \rightarrow \infty. \end{aligned}$$

Then, by independence

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq j \leq k_n} |X_{nj}| \geq \varepsilon x_n\right) &= \mathbb{P}(|X_{n1}| \geq \varepsilon x_n) + \sum_{k=2}^{k_n} \mathbb{P}\left(\max_{1 \leq j \leq k-1} |X_{nj}| < \varepsilon x_n\right) \mathbb{P}(|X_{nk}| \geq \varepsilon x_n) \\ &\geq \mathbb{P}\left(\max_{1 \leq j \leq k_n} |X_{nj}| < \varepsilon x_n\right) \sum_{k=1}^{k_n} \mathbb{P}(|X_{nk}| \geq \varepsilon x_n) \end{aligned}$$

that gives

$$\begin{aligned} \left(\sum_{j=1}^{k_n} \mathbb{P}(|X_{nj}| \geq \varepsilon x_n)\right)^2 &\leq \frac{\mathbb{P}(\max_{1 \leq j \leq k_n} |X_{nj}| \geq \varepsilon x_n)}{\mathbb{P}(\max_{1 \leq j \leq k_n} |X_{nj}| < \varepsilon x_n)} \sum_{j=1}^{k_n} \mathbb{P}(|X_{nj}| \geq \varepsilon x_n) \\ &= o\left(\sum_{j=1}^{k_n} \mathbb{P}(|X_{nj}| \geq \varepsilon x_n)\right) \text{ as } n \rightarrow \infty, \end{aligned}$$

since $x_n \rightarrow \infty$ and $\max_{1 \leq j \leq k_n} |X_{nj}|$ is stochastically bounded. \diamond

4.2 Proof of Theorem 2.2

It is convenient to normalize by the variance of partial sum and we shall consider without restricting the generality that

$$\sum_{i=1}^{k_n} c_{ni}^2 = 1 \text{ and } \max_{1 \leq i \leq k_n} c_{ni}^2 \rightarrow 0. \quad (25)$$

Then we have $\sum_{i=1}^{k_n} c_{ni}^t \leq \max_{1 \leq i \leq k_n} c_{ni}^{t-2} \rightarrow 0$ implying that $D_{nt}^{-1} \rightarrow \infty$. Moreover, the sequence $\sum_{i=1}^{k_n} c_{ni} \xi_i$ is stochastically bounded and we analyze the last three terms in Proposition 4.1. Let $x = x_n \rightarrow \infty$. By and taking into account that $x/c_{ni} \geq x \rightarrow \infty$ and h is a slowly varying function we derive for any γ fixed

$$\begin{aligned} & \sum_{i=1}^{k_n} c_{ni}^t \left(h\left(\frac{x}{c_{ni}}\right) - h\left((1+\gamma)\frac{x}{c_{ni}}\right) \right) = \\ & \sum_{i=1}^{k_n} c_{ni}^t h\left(\frac{x}{c_{ni}}\right) \left(1 - \frac{h((1+\gamma)x/c_{ni})}{h(x/c_{ni})} \right) = o\left(\sum_{i=1}^{k_n} c_{ni}^t h\left(\frac{x}{c_{ni}}\right)\right), \end{aligned}$$

implying that

$$\begin{aligned} \frac{\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq (1 \pm \eta)x)}{\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq x)} &= \frac{(1 + o_1(1)) \sum_{i=1}^{k_n} c_{ni}^t h((1 \pm \eta)x/c_{ni})}{(1 \pm \eta)^t (1 + o_2(1)) \sum_{i=1}^{k_n} c_{ni}^t h(x/c_{ni})} \rightarrow 1 \\ &\text{when } n \rightarrow \infty \text{ followed by } \eta \rightarrow 0. \end{aligned}$$

Then, we also have

$$\frac{\sum_{i=1}^{k_n} \mathbb{P}((1 - \eta)x \leq c_{ni} \xi_i < (1 + \eta)x)}{\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq x)} \rightarrow 0 \text{ as } n \rightarrow \infty \text{ and } \eta \rightarrow 0.$$

Similarly, for every $\varepsilon > 0$ fixed we have that

$$\frac{\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq \varepsilon x)}{\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq x)} = \frac{(1 + o_1(1))(1 + o_3(1))}{\varepsilon^t (1 + o_2(1))} \rightarrow \frac{1}{\varepsilon^t} \text{ as } n \rightarrow \infty.$$

and then,

$$\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq \varepsilon x) \ll \sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq x) \text{ as } n \rightarrow \infty.$$

So far, for any $\varepsilon > 0$ fixed, by letting $n \rightarrow \infty$ first and after that, passing with η to 0, we deduce by the above consideration combined with Proposition 4.1 that

$$\mathbb{P}(S_n \geq x) = \sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_i \geq x) (1 + o(1)) + \mathbb{P}(S_n^{(\varepsilon x)} \geq x) \text{ as } n \rightarrow \infty. \quad (26)$$

It remains to study the term $\mathbb{P}(S_n^{(\varepsilon x)} \geq x)$. We shall base this part of the proof on Corollary 1.7 in S. Nagaev (1979), given in the Appendix, which we apply

with $m > t$, that will be selected later. Because we assume $\mathbb{E}(\xi_0^2) = 1$, we have for all y , $B_n^2(-\infty, y) \leq 1$, and therefore, Theorem 5.1 implies:

$$\mathbb{P}(S_n^{(\varepsilon x)} \geq x) \leq \exp(-\alpha^2 x^2 / 2e^m) + (A(m; 0, \varepsilon x) / (\beta \varepsilon^{m-1} x^m))^{\beta/\varepsilon}.$$

with $\alpha = 1 - \beta = 2/(m+2)$. Then, obviously, it is enough to select $x \rightarrow \infty$ and $\varepsilon > 0$ such that

$$\exp(-\frac{\alpha^2 x^2}{2e^m}) + \left(\frac{A(m; 0, \varepsilon x)}{\beta \varepsilon^{m-1} x^m} \right)^{\beta/\varepsilon} = o \left(\sum_{i=1}^{k_n} \frac{c_{ni}^t}{x^t} h\left(\frac{x}{c_{ni}}\right) \right). \quad (27)$$

Let $x = x_n = C_n [\ln(D_{nt}^{-1})]^{1/2}$ where $C_n > e^{m/2}(m+2)/\sqrt{2}$. As we mention at the beginning of the proof $x \rightarrow \infty$.

We shall estimate each term in the left hand side of (27) separately. Because, by the definition of α we have $C_n > e^{m/2} \alpha^{-1} \sqrt{2}$, we can select $0 < \eta < 1$ such that $C_n^2 \alpha^2 / 2e^m = (1 - \eta)^{-2}$.

Taking into account the fact that for any $c > 0$ and $d > 0$ we have $y^d \exp(-cy) = o(\exp(-c(1-\eta)y))$ as $y \rightarrow \infty$, by the definition on x and η , we obtain:

$$\begin{aligned} x^{(t-2\eta)/(1-\eta)} \exp(-\frac{\alpha^2 x^2}{2e^m}) &= o(\exp(-\frac{\alpha^2 x^2}{2e^m} (1-\eta))) \\ &= o\left(\left(\sum_{i=1}^{k_n} c_{ni}^t\right)^{C_n^2 \alpha^2 (1-\eta)/2e^m}\right) = o\left(\left(\sum_{i=1}^{k_n} c_{ni}^t\right)^{(1-\eta)^{-1}}\right). \end{aligned}$$

Applying now the Hölder inequality we clearly have,

$$\sum_{i=1}^{k_n} c_{ni}^t = \sum_{i=1}^{k_n} c_{ni}^{2\eta} c_{ni}^{t-2\eta} \leq \left(\sum_{i=1}^{k_n} c_{ni}^2\right)^\eta \left(\sum_{i=1}^{k_n} c_{ni}^{(t-2\eta)/(1-\eta)}\right)^{1-\eta}. \quad (28)$$

Taking into account that $\sum_{i=1}^{k_n} c_{ni}^2 = 1$, we obtain overall

$$\exp(-\frac{\alpha^2 x^2}{2e^m}) = o \left(x^{-(t-2\eta)/(1-\eta)} \sum_{i=1}^{k_n} c_{ni}^{(t-2\eta)/(1-\eta)} \right).$$

It remains to notice that because $t > 2$, we have $(t-2\eta)/(1-\eta) > t$. Then, by combining this observation with the properties of slowly varying functions we have

$$\exp(-\frac{\alpha^2 x^2}{2e^m}) = o \left(\sum_{i=1}^{k_n} \frac{c_{ni}^t}{x^t} h\left(\frac{x}{c_{ni}}\right) \right).$$

We select ε by analyzing the second term in the left hand side of (27). Notice that by integration by parts formula, for every $z > y > 0$,

$$\begin{aligned} \mathbb{E} \xi_0^m I(0 \leq \xi_0 < z) &= \\ -z^m \mathbb{P}(\xi_0 \geq z) + m \int_0^z u^{m-1} \mathbb{P}(\xi_0 \geq u) du &\leq y^m + m \int_y^z u^{m-1} \mathbb{P}(\xi_0 \geq u) du. \end{aligned}$$

Replacing $z = \varepsilon x / c_{ni}$, taking into account condition (4), the properties of slowly varying functions, and the facts that $x / c_{ni} \rightarrow \infty$ and $m > t$, we easily obtain for y sufficiently large

$$\mathbb{E} \xi_0^m I(0 \leq c_{ni} \xi_0 < \varepsilon x) \leq y^m + 2m \int_y^{\frac{\varepsilon x}{c_{ni}}} u^{m-t-1} h(u) du = O\left(\left(\frac{x}{c_{ni}}\right)^{m-t} h\left(\frac{x}{c_{ni}}\right)\right).$$

It follows that

$$\begin{aligned} A(m; 0, \varepsilon x) &= \sum_{i=1}^{k_n} c_{ni}^m \mathbb{E} \xi_0^m I(0 \leq c_{ni} \xi_0 < \varepsilon x) \\ &\ll \sum_{i=1}^{k_n} c_{ni}^m \left(\frac{x}{c_{ni}}\right)^{m-t} h\left(\frac{x}{c_{ni}}\right) \ll x^{m-t} \sum_{i=1}^{k_n} c_{ni}^t h\left(\frac{x}{c_{ni}}\right). \end{aligned}$$

The second term, has the order

$$\left(\frac{A(m; 0, \varepsilon x)}{\beta \varepsilon^{m-1} x^m}\right)^{\beta/\varepsilon} \ll \left(\frac{x^{m-t}}{x^m} \sum_{i=1}^{k_n} c_{ni}^t h\left(\frac{x}{c_{ni}}\right)\right)^{\beta/\varepsilon} = o\left(\sum_{i=1}^{k_n} \frac{c_{ni}^t}{x^t} h\left(\frac{x}{c_{ni}}\right)\right)$$

immediately as $\beta/\varepsilon > 1$. This condition leads to the selection of ε with $0 < \varepsilon < \beta$.

Overall we obtain for any $x = C_n (\ln(\sum_{i=1}^{k_n} c_{ni}^t)^{-1})^{1/2}$ with $C_n > e^{m/2}(m+2)/\sqrt{2}$,

$$\mathbb{P}(S_n \geq x) \leq (1 + o(1)) \sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_0 \geq x) \text{ as } n \rightarrow \infty,$$

where $m > t$. Since $C_t > e^{t/2}(t+2)/\sqrt{2}$ we can select and fix $m > t$ such that $C_t > e^{m/2}(m+2)/\sqrt{2}$.

We combine this result with the lower bound to complete the proof of this Theorem. \diamond

4.3 Proof of Theorem 2.3

This result easily follows from Theorem 1 in Frolov (2005) when moments strictly larger than 2 are available. This Theorem is given for convenience in the Appendix (Theorem 5.2). Because we assume the existence of moments of order p , we have

$$\begin{aligned} \Lambda_n(u, s, \epsilon) &\leq \frac{u}{\sigma_n^2} \sum_{j=1}^{k_n} c_{nj}^2 \mathbb{E} \xi_0^2 I(|c_{nj} \xi_0| > \epsilon \sigma_n / s) \\ &\leq \frac{us^{p-2}}{\sigma_n^p \epsilon^{p-2}} \sum_{j=1}^{k_n} |c_{nj}|^p \mathbb{E} |\xi_0|^p = \epsilon^{2-p} us^{p-2} L_{np}. \end{aligned}$$

where $L_{np} = \sigma_n^{-p} \sum_{j=1}^{k_n} |c_{nj}|^p \mathbb{E}|\xi_0|^p$. Then, for $x^2 \leq (2 \ln(1/L_{np}))$,

$$\Lambda_n(x^4, x^5, \epsilon) \leq \epsilon^{2-p} x^{4+5(p-2)} L_{np} \leq \epsilon^{2-p} L_{np} (2 \ln(1/L_{np}))^{(5p-6)/2}.$$

The proof is immediate from Theorem 5.2. Just notice that by (10)

$$L_{np} \leq \frac{\max_{1 \leq j \leq k_n} |c_{nj}|^{p-2}}{\sigma_n^{p-2}} \mathbb{E}|\xi_0|^p \rightarrow 0.$$

Then $x^2 - 2 \ln(L_{np}^{-1}) - (p-1) \ln \ln(L_{np}^{-1}) \rightarrow -\infty$ provided $x^2 \leq 2 \ln(L_{np}^{-1}) + p/2 \ln \ln(L_{np}^{-1})$. It remains to notice that for n sufficiently large $x^2 \leq 2 \ln(D_{np}^{-1}) + \ln \ln(D_{np}^{-1})$ and the result follows. \diamond

4.4 Proof of Theorem 2.1

For simplicity we normalize by the variance and assume (25). Without restricting the generality we assume $2 < p < t$. We start from inequality (26) and apply Proposition 5.1 to the second term in the right hand side. We obtain for any $\varepsilon > 0$ and $x^2 \leq c_\varepsilon \ln(D_{np}^{-1})$ with $c_\varepsilon < 1/\varepsilon$ and for all n sufficiently large $\mathbb{P}(S_n^{(\varepsilon x)} \geq x) = (1 - \Phi(x))(1 + o(1))$. We notice now that by (28) applied with $\eta = (t-p)/(t-2)$ and simple considerations,

$$D_{nt} \leq D_{np} \leq (D_{nt})^{(p-2)/(t-2)}. \quad (29)$$

So far, by using this last relation, we practically showed that (11) holds for $0 < x \leq C[\ln(D_{nt}^{-1})]^{1/2}$ with C an arbitrary positive number. On the other hand, because $(1 - \Phi(x)) \leq x^{-1} \exp(-x^2/2)$, by Theorem 2.2 and by the arguments leading to the proof of relation (27), there is a constant $c > 0$ such that for $x > c[\ln(D_{nt}^{-1})]^{1/2}$, we simultaneously have

$$\mathbb{P}(S_n \geq x) = (1 + o(1)) \sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_0 \geq x)$$

and

$$(1 - \Phi(x)) = o\left(\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_0 \geq x)\right).$$

Then (11) holds for all $x > 0$ since C is arbitrarily large and can be selected such that $c < C$. \diamond

4.5 Proof of Corollary 2.1

The ideas involved in the proof of this corollary already appeared in the previous proofs, so we shall mention only the changes. We start from (11). To prove (12) we have to show that

$$1 - \Phi(x) = o\left(\sum_{i=1}^{k_n} \mathbb{P}(c_{ni} \xi_0 \geq x \sigma_n)\right)$$

for $x \geq a(\ln D_{nt}^{-1})^{1/2}$ with $a > 2^{1/2}$. First we shall use the relation $1 - \Phi(x) \leq x^{-1} \exp(-x^2/2)$. Then, we adapt the proof we used to establish the first part of (27), when we compared $\exp(-\alpha^2 x^2/2e^m)$ to $\sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \geq x\sigma_n)$. The main difference is that now we take $m = 0$ and $\alpha = 1$.

For the proof of (13), we use the inequality $1 - \Phi(x) \geq (1+x)^{-1} \exp(-x^2/2)$. By (4) and (29) we have for every $0 < \varepsilon < t - 2$,

$$\sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \geq x\sigma_n) \ll \sum_{i=1}^{k_n} \frac{c_{ni}^{t-\varepsilon}}{x^{t-\varepsilon}} \ll \frac{1}{x^{t-\varepsilon}} (D_{nt})^{(t-2-\varepsilon)/(t-2)}.$$

Then, it is easy to see that, because ε can be made arbitrarily small, for $1 < x \leq b(\ln D_{nt}^{-1})^{1/2}$ with $b < 2^{1/2}$ we have

$$\sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \geq x\sigma_n) = o(1 - \Phi(x)).$$

When $0 < x \leq 1$ we apply Theorem 2.3. \diamond

4.6 Proof of Corollary 2.5

This Corollary follows from Corollary 2.4 via Lemma 5.1 in the Appendix. It remains to give an explicit form of the intervals moderate deviation and large deviation boundaries. For proving the large deviation part of this corollary we have to analyze the condition on x from part (ii) of Corollary 2.4, namely $x > a(\ln D_{nt}^{-1})^{1/2}$ with $a = \sqrt{2}$. By Lemma 5.1

$$B_{n2} = \sum_i b_{ni}^2 \sim c_r n^{3-2r} l^2(n)$$

and

$$C_1 l^t(n) n^{(1-r)t+1} \leq B_{nt} = \sum_{j=1}^{\infty} b_{n,j}^t \leq C_2 l^t(n) n^{(1-r)t+1}.$$

Then, for certain constants K_1 and K_2 and because $D_{nt}^{-1} = B_{n2}^{t/2}/B_{nt}$, we have for n sufficiently large

$$K_1 + \ln n^{(t-2)/2} \leq \ln D_{nt}^{-1} \leq K_2 + \ln n^{(t-2)/2}.$$

So, the asymptotic result (12) holds for $x \geq c_1(\ln n)^{1/2}$ where $c_1 > (t-2)^{1/2}$. Furthermore, (13) holds for $0 < x \leq c_2(\ln n)^{1/2}$ where $c_2 < (t-2)^{1/2}$. \diamond

4.7 Proof of Theorem 2.4

Without restricting the generality we assume $\kappa > 0$, since similar computations can be done when $\kappa < 0$. Let $A_n = \sum_{i=n}^{\infty} a_i^2$. Using the argument of Theorem 5 in Wu (2006), under Condition B, we have

$$\|\mathcal{P}_0(K(X_n) - \kappa X_n)\|_q = O(\theta_n), \text{ where } \theta_n = |a_n|^{p/q} + |a_n| A_n^{1/2}.$$

Let $\theta_i = 0$ if $i \leq 0$ and $\Theta_n = \sum_{i=1}^n \theta_i$. Then by Theorem 1 in Wu (2007), there exists a constant $B_q \geq 1$ such that

$$\frac{\|S_{n,1}\|_q^2}{B_q^2} \leq \sum_{i \in \mathbb{Z}} (\Theta_{n+i} - \Theta_i)^2 \leq 2n\Theta_{2n}^2 + \sum_{i=n+1}^{\infty} (\Theta_{n+i} - \Theta_i)^2. \quad (30)$$

By Karamata's theorem, $A_n \sim (2r-1)^{-1}n^{1-2r}l(n)^2$, and if $i > n$, $\Theta_{n+i} - \Theta_i = O(n\theta_i)$ and $\sum_{i=n+1}^{\infty} \theta_i^2 = O(n\theta_n^2)$. Let $\ell(\cdot)$ be a slowly varying function and $\beta \in \mathbb{R}$. Again by Karamata's theorem, there exists another slowly varying function $\ell_0(\cdot)$ such that $\sum_{i=1}^n i^{-\beta}\ell(i) = O(1+n^{1-\beta})\ell_0(n)$. Hence by (30), there exists a slowly varying function $\ell_1(\cdot)$ such that

$$\|S_{n,1}\|_q = O(\sqrt{n})(1 + n^{1-rp/q} + n^{1-r+(1-2r)/2})\ell_1(n). \quad (31)$$

For $n \geq 3$ let $g_n = (\ln n)^{-1}$. Then

$$\mathbb{P}(S_n \geq (x + g_n)\sigma_n) - \mathbb{P}(H_n \geq \kappa x \sigma_n) \leq \mathbb{P}(|S_{n,1}| \geq \kappa g_n \sigma_n). \quad (32)$$

Since $x^2 \leq c \ln n$ and $g_n = (\ln n)^{-1}$, we have that $1 - \Phi(x \pm g_n) \sim 1 - \Phi(x)$. Hence by Corollary 2.5, (21) follows from (32) in view of

$$\begin{aligned} \mathbb{P}(|S_{n,1}| \geq \kappa g_n \sigma_n) &\leq \frac{\|S_{n,1}\|_q^q}{|\kappa|^q g_n^q \sigma_n^q} = \frac{O(\sqrt{n}^q)(1 + n^{q-rp} + n^{(3/2-2r)q})\ell_1^q(n)}{g_n^q (n^{3/2-r}l(n))^q} \\ &= n^{-p\rho(r)} \frac{\ell_1^q(n)}{g_n^q l^q(n)} = \frac{o(n^{-c/2})}{\ln n} = o(xe^{-x/2}) = o[1 - \Phi(x)], \end{aligned} \quad (33)$$

since $c/2 < p\rho(r)$. Here we note that $\ell_1(n)/(g_n l(n))$ is also slowly varying in n and $x \leq c \ln n$. By (31) and (33), it is easily seen that the normalizing constant $\kappa \sigma_n$ can be replaced by $\sqrt{\text{var}(H_n)}$. The proof of the upper bound is similar and it is left to the reader. \diamond

5 Appendix

The following Theorem is a slight reformulation of Fuk–Nagaev inequality (see Corollary 1.7, S. Nagaev, 1979):

Theorem 5.1 *Let Y_1, Y_2, \dots, Y_n be independent random variables and $m \geq 2$. Suppose $\mathbb{E}Y_i = 0$, $i = 1, \dots, n$, $\beta = m/(m+2)$, and $\alpha = 1 - \beta = 2/(m+2)$. For $y > 0$, define $Y^{(y)} = Y_i I(Y_i \leq y)$, $A_n(m; 0, y) := \sum_{i=1}^n \mathbb{E}[Y_i^m I(0 < Y_i < y)]$ and $B_n^2(-\infty, y) := \sum_{i=1}^n \mathbb{E}[Y_i^2 I(Y_i < y)]$. Then for any $x > 0$ and $y > 0$*

$$\mathbb{P}\left(\sum_{i=1}^n Y_i^{(y)} \geq x\right) \leq \exp\left(-\frac{\alpha^2 x^2}{2e^m B_n^2(-\infty, y)}\right) + \left(\frac{A_n(m; 0, y)}{\beta x y^{m-1}}\right)^{\beta x/y}.$$

We shall also use the following result which is an immediate consequence of Theorem 1.1 in Frolov (2005).

Theorem 5.2 Let $(X_{nj})_{1 \leq j \leq k_n}$ be an array of row-wise independent centered random variables. Let $p > 2$ and denote $S_n = \sum_{j=1}^{k_n} X_{nj}$, $\sigma_n^2 = \sum_{j=1}^{k_n} \mathbb{E}X_{nj}^2 \rightarrow \infty$, $M_{np} = \sum_{j=1}^{k_n} \mathbb{E}X_{nj}^p I(X_{nj} \geq 0) < \infty$, $L_{np} = \sigma_n^{-p} M_{np}$ and denote

$$\Lambda_n(u, s, \epsilon) = \frac{u}{\sigma_n^2} \sum_{j=1}^{k_n} \mathbb{E}X_{nj}^2 I(X_{nj} \leq -\epsilon \sigma_n / s).$$

Furthermore, assume $L_{np} \rightarrow 0$ and $\Lambda_n(x^4, x^5, \epsilon) \rightarrow 0$ for any $\epsilon > 0$. Then if $x \geq 0$ and $x^2 - 2 \ln(L_{nt}^{-1}) - (t-1) \ln \ln(L_{nt}^{-1}) \rightarrow -\infty$, we have

$$\mathbb{P}(S_n \geq x \sigma_n) = (1 - \Phi(x))(1 + o(1)).$$

For truncated random variables by following the proof of Theorem 1.1 in Frolov (2005) we can present his relation (3.17) as a proposition.

Proposition 5.1 Assume the conditions in Theorem 5.2 are satisfied. Define

$$X'_{nk} = X_{nk} I(X_{nk} \leq \varepsilon x \sigma_n) \text{ and } S'_n = \sum_{j=1}^{k_n} X'_{nj}.$$

Fix $\varepsilon > 0$. Then if $x^2 \leq c \ln(L_{np}^{-1})$ with $c < 1/\varepsilon$, for all n sufficiently large we have

$$\mathbb{P}(S'_n \geq x \sigma_n) = (1 - \Phi(x))(1 + o(1)).$$

The following facts about the series are going to be used to analyze a class of linear processes:

Lemma 5.1 Assume $a_i = l(i)i^{-r}$ with $1/2 < r < 1$. Let $b_j := b_{nj} := \sum_{i=1}^j a_i$ if $1 \leq j \leq n$ and $b_{nj} := \sum_{i=j-n+1}^j a_i$ if $j > n$. Then, for two positive constants C_1 and C_2 , we have

$$C_1(l^t(n)n^{(1-r)t+1}) \leq \sum_{j=1}^{\infty} b_{nj}^t \leq C_2(l^t(n)n^{(1-r)t+1}),$$

for any $t \geq 2$. In the case $t = 2$, $\sum_{j=1}^{\infty} b_{nj}^2 = c_r n^{3-2r} l^2(n)$ with

$$c_r = \left\{ \int_0^{\infty} [x^{1-r} - \max(x-1, 0)^{1-r}]^2 dx \right\} / (1-r)^2.$$

Proof. It is easy to see that $b_{nj} \ll j^{1-r} l(j)$ for $j \leq 2n$ and $b_{nj} \ll n(j-n)^{-r} l(j)$ for $j > 2n$ from the Karamata theorem (see part 1 of Lemma 5.4 in Peligrad and Sang (2010)). Therefore,

$$\begin{aligned} \sum_{j=1}^{\infty} b_{nj}^t &= \sum_{j=1}^{2n} b_{nj}^t + \sum_{j=2n+1}^{\infty} b_{nj}^t \\ &\ll \sum_{j=1}^{2n} j^{(1-r)t} l^t(j) + \sum_{j=2n+1}^{\infty} n^t (j-n)^{-rt} l^t(j) = O(l^t(n)n^{(1-r)t+1}). \end{aligned}$$

The proof in the other direction is similar. The result of case $t = 2$ is well known. See for instance Theorem 2 in Wu and Min (2005). \diamond

References

- [1] Acerbi, C. and D. Tasche (2002). On the coherence of expected shortfall. *J. Banking and Finance* **26** 1487–1503.
- [2] Beran, J. (1994). Statistics for long-memory processes. *Monographs on Statistics and Applied Probability* **61**, Chapman and Hall, New York.
- [3] Bingham, N. H., C. M. Goldie and J. L. Teugels (1987). *Regular Variation*. Cambridge, UK: Cambridge University Press.
- [4] Chen, X. (2001). Moderate deviations for Markovian occupation times. *Stochastic Process. Appl.* **94**, 51–70.
- [5] de la Peña, V. and E. Giné (1999). *Decoupling. From dependence to independence*. Springer.
- [6] Doukhan, P., G. Oppenheim and M. S. Taqqu (editors) (2003). *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston.
- [7] Frolov, A. N. (2005). On probabilities of moderate deviations of sums for independent random variables. *Journal of Mathematical Sciences* **127**, 1787–1796.
- [8] Ghosh, M. (1974). Probabilities of moderate deviations under m -dependence. *Canad. J. Statist.* **2**, 157–168.
- [9] Grama, I. G. (1997). On moderate deviations for martingales. *Ann. Probab.* **25**, 152–183.
- [10] Grama, I. G. and E. Haeusler (2006). An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *J. Theoret. Probab* **19**, 1–44.
- [11] Granger, C. W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal* **1** 15–29.
- [12] Hall, P. (1992). Convergence rates in the central limit theorem for means of autoregressive and moving average sequences. *Stochastic Process. Appl.* **43**, 115–131.
- [13] Ho, H. C. and T. Hsing (1997). Limit theorems for functionals of moving average. *Ann. Probab.* **25**, 1636–1669.
- [14] Holton, Glyn (2003). *Value-at-Risk: Theory and Practice*. Academic Press.
- [15] Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–176.
- [16] Hurst, Simon, The Characteristic Function of the Student-t Distribution, Financial Mathematics Research Report No. FMRR006-95, Statistics Research Report No. SRR044-95

- [17] Jorion, Philippe (2006). *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd ed.). McGraw-Hill.
- [18] McNeil, Alexander, Frey, Rüdiger and Embrechts, Paul (2005). *Quantitative Risk Management: Concepts Techniques and Tools*. Princeton University Press.
- [19] Mikosch, T. and A. V. Nagaev (1998). Large Deviations of Heavy-Tailed Sums with Applications in Insurance. *Extremes* **1:1**, 81–110.
- [20] Nagaev, A. V. (1969). Limit theorems for large deviations where Cramér’s conditions are violated (in Russian). *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk* **6**, 17–22.
- [21] Nagaev, S. V. (1965). Some limit theorems for large deviations. *Teor. Veroyatn. Primen.* **10**, 231–254.
- [22] Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745–789.
- [23] Peligrad, M and H. Sang (2010). Asymptotic properties of self-normalized linear processes with long memory. submitted
- [24] Peligrad, M. and S. Utev (1997). Central limit theorem for linear processes. *Ann. Probab.* **25** 443–456.
- [25] Robinson, P. M. (1997). Large-sample inference for non parametric regression with dependent errors. *Ann. Statist.* **25**, 2054–2083.
- [26] Robinson, P. M. (2003). *Time series with long memory*, Oxford University Press
- [27] Rozovski, L. V. (1993). Probabilities of large deviations on the whole axis. *Theory Probab. Appl.* **38**, 53–79.
- [28] Rubin, H. and J. Sethuraman (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27**, 325–346.
- [29] Slastnikov, A. D. (1978). Limit theorems for probabilities of moderate deviations *Teor. Veroyatn. Primen.* **24**, 340–357
- [30] Wu, W. B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory* **22**, 1–14.
- [31] Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Annals of Probability* **35**, 2294–2320.
- [32] Wu, W. B. and W. Min (2005). On Linear Processes with Dependent Innovations. *Stochastic Processes and their Applications* **115**, 939–958.
- [33] Wu, W. B. and Z. Zhao (2008). Moderate deviations for stationary processes. *Statistica Sinica* **18**, 769–782.